

TEST DEVELOPMENT PROJECT (TDP)

Rohmani Nur Indah

NIM: 100221608049

A Project Report of

Assessment Instrument Development

Supervised by

Prof. Dr. M. Soenardi Djiwandono



DOCTORATE PROGRAM OF ENGLISH LANGUAGE TEACHING

STATE UNIVERSITY OF MALANG

2010

TABLE OF CONTENT

| | |
|--|-----------|
| A. TEST DEVELOPMENT PROJECT (TDP) | 3 |
| 1. Preparation | 3 |
| 2. Test Plan | 3 |
| 3. Test Development | 6 |
| 4. Test Validation | 7 |
| 5. Test Try-out | 7 |
| 6. Project Report | 8 |
| B. ATTACHMENTS | 10 |
| 1. Attachment 2.3: What to Test & How to Test | 10 |
| 2. Attachment 2.7: Table of Specification | 11 |
| 3. Attachment 2.10.1: Reading Test Item Identification | 12 |
| 4. Attachment 2.10.2: Table of answers, location in text, spec. obj & question | 13 |
| 5. Attachment 3.2: Scoring Guide – Setting up Criteria | 14 |
| 6. Attachment 3.4: Grading of Scores | 15 |
| 7. Attachment 4.1.1: Complete Test Items | 16 |
| 8. Attachment 4.1.2: Revised Test Items | 17 |
| 9. Attachment 4.3: The Reliability Formula | 20 |
| 10. Attachment 5.1: Summary of Notes of Tryout Test | 21 |
| 11. Attachment 5.2.1: Table of Scores | 22 |
| 12. Attachment 5.2.2: Simple Statistics | 25 |
| 13. Attachment 5.2.3: Test Item Analysis | 28 |
| 14. Attachment 5.3: The Grades | 29 |
| 15. Attachment 5.4: Some of The Problems & How to Improve | 30 |
| C. REVIEW OF THE RELATED THEORIES | 31 |
| 1. Introduction | 31 |
| 2. Reading | 31 |
| Definition of reading | 31 |
| Nature of reading | 32 |
| Types of reading skills | 34 |
| Reading model and stages | 35 |
| 3. Testing Reading | 37 |
| What to test in reading | 38 |
| Different ways of testing reading | 38 |
| Reading test format | 39 |
| Reading test item development | 39 |
| 4. Reading Test Analysis | 41 |
| Scoring of reading test | 41 |
| Grading of reading test | 41 |
| The validity of reading test | 42 |
| The reliability reading test | 43 |
| 5. Lessons Learned in Reading Test Development | 45 |
| 6. References | 46 |

A. TEST DEVELOPMENT PLAN (TDP)

Draft: 6

| STAGES | DETAILS/DESCRIPTION (in the form of brief notes) | ATTACHMENT (numbers labels) |
|--|--|--------------------------------|
| (1) PREPARATION | | |
| 1.1 Name of school | State University of Malang | - |
| 1.2 Faculty Department | Letters English Language Teaching | - |
| 1.3 Subject Grade Level | Reading II (Interpretive and Affective Reading) S1 – second year Pre-advanced Reading I: Literal reading (semester 2) Reading II: Interpretive Reading (semester 3) Reading III: Critical Reading (semester 4) <u>Reference:</u> Syllabus UM, 2010 Malang: UM Press | - |
| 1.4 Number of Students Teacher's name | 30 students Nur Hayati, S.Pd, M.Ed There are five offerings of Reading II with the same number of students taught by different teachers. Miss Nur Hayati teaches one offering. | - |
| (2) TEST PLAN | | |
| 2.1 General objective of the course | <u>Quotation:</u> 1) Interpret non-fictional texts of pre-advanced level, with a focus on inferring and applying 2) Appreciate literary works of pre-advanced level with a focus on inferring and applying <u>Reference:</u> Syllabus UM, 2010, Malang: UM Press <u>Comment:</u> not appropriate, too specific. | - |

| | | |
|---|---|------------------------------------|
| | <p><u>Proposed:</u> Students can understand English written texts.</p> <p><u>Justification:</u> Reading comprehension is basically the ability of understanding ideas in written text (Djiwandono, 1996).</p> | |
| 2.2 Specific objective of the course | <p><u>Quotation:</u> 1.1) Understand implicitly stated information 1.2) Interpret cohesive devices in a text 1.3) Recognize discourse markers 1.4) Identify the organization and development of a text 1.5) Identify types of writing 2.1) Appreciate short stories</p> <p><u>Reference:</u> Syllabus UM, 2010</p> <p><u>Comment:</u> not appropriate, not clear.</p> <p><u>Proposed:</u> Students can (1) understand explicitly stated information; (2) understand implicitly stated information of English written texts.</p> | - |
| 2.3 Object of test | Reading non-fictional text | Att. 2.3: WHAT & HOW TO TEST |
| 2.4 General objective of test | <p><i>Not available</i></p> <p><u>Proposed:</u> To test students' ability to understand English written text.</p> <p><u>Justification:</u> It refers to 2.1. The proposed general objective of the course</p> | - |
| 2.5 Specific objective of test | <p><i>Not available</i></p> <p><u>Proposed:</u> To test students' ability to understand explicitly stated information and implicitly stated information from English written text.</p> | - |

| | | |
|--------------------------------------|--|---|
| | <u>Justification:</u> It refers to 2.2 The proposed specific objective of the course. | |
| 2.6 Kind and format of test | Wh-questions. | - |
| 2.7 Table of specification | The test is to assess the understanding of the explicitly stated information in 5 wh-questions (70%) and implicitly stated information in 5 wh-questions (30%) | Att.2.7: TOS |
| 2.8 Source of test materials | <u>Textbook used:</u> <i>Interpretive and Affective Reading</i> . M. Djoehana D. Oka & Sri Andriani. Depdiknas UM. Fak. Sastra Jur. Sastra Inggris 2008. <u>Justification:</u> The textbook is completed with reading strategies and reading practices | - |
| 2.9 Text used for the test | The Sporting Spirit By George Orwell (<i>Interpretive and Affective Reading</i> . M. Djoehana D. Oka & Sri Andriani. Depdiknas UM. Fak. Sastra Jur. Sastra Inggris 2008.) <u>Description:</u> The text from the textbook is not totally unseen which means that students may have been familiar with the vocabulary used in the text. Content: popular topic on sport Language: easy to understand Number of words: 235 words. | - |
| 2.10 Test item identification | One text and 10 wh-questions | Att.2.10.1: TEST ITEM IDENTIFICATION: READING Att.2.10.2: TABLE OF ANSWERS, LOCATION IN TEXT, SPEC.OBJ, AND READING TEST ITEMS |

| (3) TEST DEVELOPMENT | | |
|----------------------------|---|---|
| 3.1 Writing of test | <p>The test items are developed in 10 wh-questions, 7 questions on explicit information and 3 questions on implicit information.</p> <p>Time allocation: 60 minutes</p> <p>Direction:</p> <ol style="list-style-type: none"> 1. Read the following passage carefully 2. Answer the questions based on the stated or implied information in the text. Write the answer in brief (not necessarily in the form of full sentence) <p>E.g.: What is the author's mainpoint in the text? (<i>Answer: the sporting spirit</i>)</p> | |
| 3.2 Scoring guide | <p>The criterion of scoring each test item is</p> <ol style="list-style-type: none"> 1. Content: the correctness of the ability to pick up explicitly or implicitly stated information from the text. 2. Language: the correct use of grammar and vocabulary | Att. 3.2: SCORING GUIDE: SET UP CRITERIA |
| 3.3 Scoring | <p>Who: The scorers are the test developer and the lecturer.</p> <p>How: By confirming to the scoring guide and the minimum acceptable level. If there is big difference (≥ 3) for each score, a third scorer is needed. Then the scores from all scorers are added and divided by the number of the scorers to give average raw score.</p> | - |
| 3.4 Grading | <p>CR</p> <ol style="list-style-type: none"> 1. Deciding criteria: Content Appropriacy 2. Defining description of each criterion: a) correct, b) relatively correct, c) incorrect 3. Giving score to each part of the descriptor ranging from 3-1 4. Determining the minimum acceptable level that is giving relatively correct answer 5. Choosing the minimum acceptable score by referring to the criteria 6. Deciding the minimum passing grade (C) 7. Determining the score range and grading | Att. 3.4: GRADING: CONVERSION OF SCORES |

| (4) TEST VALIDATION | | |
|--------------------------------------|---|--|
| 4.1 Review of the entire test | Galuh Nur Rohmah, M.Ed., M.Pd did the review. She is the head of English Department of UIN Malang who has been teaching reading for more than 10 years. Her feedbacks on the test draft are attached. | Att. 4.1.1: COMPLETE TEST DRAFT Att.4.1.2: COMPLETE REVISED DRAFT |
| 4.2 Validity | Content validity. It relies on careful analysis of the language being used in the test. The content represents the materials in the course objective and measure the ability to understand explicit and implicit information of the reading text. | - |
| 4.3 Reliability | Inter-rater reliability. Two test-scorers, the test developer and the teacher check the answers based on scoring guide (att.3.2). If there is a big difference (≥ 3) for each score, a third scorer is invited to decide the average final score. The reliability is calculated by using Pearson-Product Moment . The formula is adapted from Djiwandono, 2008: p.173. It is supposed to compare the scores not between the scores of two different test, but between the scores of two different raters. | Att. 4.3: RELIABILITY FORMULA |
| (5) TEST TRYOUT | | |
| 5.1 Administration of Tryout | Day/date/time: Friday/November 19th, 2010 at 14.00 Place: room D8-202 Student number: out of 30 students, 25 attended the test Test administrator: The teacher lead the test, the test developer monitored the process Important incident: <ul style="list-style-type: none"> • A student asked about no.4 concerning whose suggestion is being asked and the test developer informed that it is the author's suggestion • Most students finished the test earlier. The last student submitted the test sheet at 14.45 | Att. 5.1: SUMMARY OF NOTES OF TRYOUT TEST |
| 5.2 Scoring | The scoring was done by the test developer and the teacher based on the steps in 3.3. Because there is no score exceeding the acceptable difference of scores, there is no need to invite | Att. 5.2.1: TABLE OF SCORES Att. 5.2.2: |

| | | |
|--|---|---|
| | <p>another rater.</p> <p>The scores are arranged based on the size of scores from the highest score (57) to the lowest score (33). The scores of each scorer are compared to find the average score to be put in att. 5.2.1 Table of scores.</p> <p>The scores are reviewed in terms of the central tendency and dispersion in att. 5.2.2 Simple statistics.</p> <p>The analysis of the quality of each test item is presented in att. 5.2.3</p> <p>Interpretation of the scores:</p> <ul style="list-style-type: none"> • As the mean is 46 and the possible highest score is 60, it means that the test is not too difficult. • The range is 24 out of the possible range 60 which means that the dispersion is not too far. • Based on level of difficulty, the test items are various: 5 easy, 2 moderate, 3 difficult • Reliability is very high (0.99) meaning that both scorers have higher agreement in scoring • Based on item analysis, 8 items are good and 2 items are bad. | <p>SIMPLE STATISTICS Att. 5.2.3: TEST ITEM ANALYSIS</p> |
| 5.3 Grading | <p>The grading is done by referring to 3.4 by using the raw score obtained from the tryout test.</p> <p>5 students got A, 14 students got B, 2 students got C and 4 students got D which means 84% of the class pass and 16% not pass</p> | Att. 5.3: GRADES |
| 5.4 Analysis of Tryout Test Results | <p>What to improve: based on the question raised by the testee, no. 4 is revised. The question needs clear wording as the text shows 2 views: the author and people's assumption.</p> <p>How to improve: Changing no. 4 into: What is the author's suggestion for the nations to avoid battle?</p> | Att. 5.4: TABLE OF SUMMARY OF PROBLEMS AND HOW TO IMPROVE |
| (6) PROJECT REPORT | | |
| Content: | <p>A. Completed TDP</p> <p>B. Complete attachment</p> <p>C. Review of Theories on:</p> <ol style="list-style-type: none"> 1. Reading <ol style="list-style-type: none"> a. Definition of reading b. The Nature of Reading c. Types of reading skills | |

| | |
|--|--|
| | <ul style="list-style-type: none"> d. Reading models and stages 2. Testing reading <ul style="list-style-type: none"> a. What to test in reading b. Different ways of testing reading c. Reading test format d. Reading test item development 3. Test analysis <ul style="list-style-type: none"> a. Scoring reading test b. Grading reading test c. The validity of reading test d. The realibility of reading test 4. Lessons learned in developing reading test 5. List of reference |
|--|--|

B. ATTACHMENTS

Attachment 2.3:

WHAT TO TEST AND HOW TO TEST

| OBJECT OF TEST | OBJECTIVES OF TEST | KIND OF PROFICIENCY | SUBJECTIVE TEST | OBJECTIVE TEST |
|----------------|---------------------------------------|---------------------|-----------------|----------------|
| LANGUAGE SKILL | | | | |
| READING | To test understanding of written text | Passive | Wh-questions | - |

TABLE OF SPECIFICATION (TOS)

SPECIFICATION OF WHAT TO TEST AND THEIR PROPORTION

| OBJECT OF TEST | OBJECTIVES OF TEST | SPECIFICATION OF OBJECTIVES | PROPORTION | TEST FORMAT |
|-----------------------------|---|--|---|---|
| Reading non-fictional text. | To assess students' ability to understand English written text. | To test students' ability to understand explicitly stated information and implicitly stated information from English written text. | <p>To test the ability to identify implication from explicitly stated information which is given in 7 wh-questions (70%)</p> <p>To test the ability to make inference from implicitly stated information which is given in 3 wh-questions (30%)</p> | Subjective test consisting 10 wh-questions. |

READING TEST ITEM IDENTIFICATION

THE SPORTING SPIRIT

1 I am always (2,E) amazed when I hear people saying that (1,E) sport creates goodwill
between the nations, and that if only the common people of the world could (4,I) meet one
3 another at football or cricket, they would have no inclination to meet on the battlefield. Even if
one didn't know from concrete examples (the 1936 (3,E) Olympic Games, for instance) that
5 international sporting contests lead to hatred, one could deduce it from general principles.

Nearly all the sports practiced nowadays are (5,E) competitive. You play to win, and the
7 game has little meaning unless you do your utmost to win. On the village green, where you pick
up sides and (7,I) no feeling of local patriotism is involved, it is possible to play simply for the
9 (6,E) fun and exercise; but as soon as the question of (8,E) prestige arises, as soon as you feel
that you and some larger unit will be disgraced if you lose, the most savage combative instincts
11 are aroused. Anyone who has played even in a school football match knows this. At the (9,E)
international level sport is frankly mimic warfare. But the significant thing is not the behaviour of
13 the players but (10,I) the attitude of the spectators and behind the spectators, of the nations
who work themselves into furies over these absurd contests, and seriously believe—at any rate
15 for short periods—that running, jumping and kicking a ball are tests of national virtue.

No. of words: 234

Author: George Orwell

Source: *Interpretive and Affective Reading*. M. Djoehana D. Oka & Sri Andriani.

Depdiknas UM. Fak. Sastra Jur. Sastra Inggris 2008.

Level: Pre-advanced.

TABLE OF ANSWERS, LOCATION IN TEXT, SPEC.OBJ., AND TEST QUESTIONS

| TEST ITEM NO. | ANSWER IN TEXT | SPECIFIC OBJ. | LOCATION IN TEXT | THE QUESTIONS |
|---------------|--|---------------|------------------|--|
| 1. | Sport | E | Line 1 | According to some people, what helps establish goodwill between nations? |
| 2. | Amazed | E | Line 1 | How is the author's feeling when he heard about the benefit of sport? |
| 3. | Olympic Games | E | Line 4 | What is the example of international sport contest given by the author? |
| 4. | Meet at a match | I | Line 2 | What is the author's suggestion for the nations to avoid battle? |
| 5. | Competitive | E | Line 6 | What is the main feature of sport today? |
| 6. | Fun and exercise | E | Line 9 | What is the common purpose of doing sport at villages? |
| 7. | Not just for fun but for prestige/local patriotism | I | Line 8 | How is doing sport at villages different from doing sport in larger units? |
| 8. | Prestige | E | Line 9 | What might make us feel disgraced when we lose? |
| 9. | International level sport | E | Line 11 | What may be similar to combat or war? |
| 10. | Bad attitude of the spectators | I | Line 13 | What does the author suggest us to be aware of in any sport match? |

E: Explicit information, I: Implicit information

SCORING GUIDE – SETTING UP CRITERIA

| OBJECT OF TEST | OBJECTIVES OF TEST | SPECIFIC OBJ. OF TEST | DESCRIPTORS | SCORE | MAL |
|---|------------------------------------|---|--|-------|-----|
| Reading | Ability to understand reading text | 1. Ability to understand explicitly stated information. | <u>Content:</u> | | |
| | | | 1. Correctly able to pick-up explicitly stated information. | 3 | |
| | | | 2. Almost correctly able to pick-up explicitly stated information. | 2 | 2 |
| | | | 3. Incorrectly able to pick-up explicitly stated information | 1 | |
| | | | <u>Language:</u> | | |
| | | | 1. Correct use of grammar and vocabulary | 3 | |
| | | | 2. Almost correct use of grammar and vocabulary | 2 | 2 |
| | | | 3. Incorrect use of grammar and vocabulary | 1 | |
| | | | Tot: 4 | | |
| | | 2. Ability to understand implicitly stated information. | <u>Content:</u> | | |
| | | | 1. Correctly able to pick-up implicitly stated information. | 3 | |
| | | | 2. Almost correctly able to pick-up implicitly stated information. | 2 | 2 |
| | | | 3. Incorrectly able to pick-up implicitly stated information | 1 | |
| | | | <u>Language:</u> | | |
| 1. Correct use of grammar and vocabulary | 3 | | | | |
| 2. Almost correct use of grammar and vocabulary | 2 | | 2 | | |
| 3. Incorrect use of grammar and vocabulary | 1 | | | | |
| Tot: 4 | | | | | |

MAL: Minimum Acceptable Level

Adapted from:Djiwandono, M.S. 2008. *Tes Bahasa: Pegangan bagi Pengajar Bahasa. Jakarta: Indeks.*

Attachment 3.4:

GRADING: CONVERSION OF SCORES

| LEVEL OF ACHIEVEMENT | TOTAL SCORES | QUALIFICATION | GRADE |
|----------------------|--|---------------|-------|
| = CRITERIA | 4 (total of MAL) X 10 (number of items) = 40 | PASS | C |
| ≥ CRITERIA | 51 – 60 | PASS | A |
| ≥ CRITERIA | 41 – 50 | PASS | B |
| ≤ CRITERIA | 20 – 39 | NOT PASS | D |
| ≤ CRITERIA | 0 – 19 | NOT PASS | E |

Adapted from:

Djiwandono, M.S. 2008. *Tes Bahasa: Pegangan bagi Pengajar Bahasa. Jakarta: Indeks.*

Pages: 251

Attachment 4.1.1: COMPLETE TEST ITEMS

TEST OF READING II (Interpretive and Affective Reading)

Directions:

1. Read the following passage.
2. Answer the questions in brief (not more than 3 words) or simply by giving a Noun Phrase as in the following example:
What is the text mainly discuss? (Answer: sporting spirit)

THE SPORTING SPIRIT

1 I am always amazed when I hear people saying that sport creates goodwill between the
nations, and that if only the common peoples of the world could meet one another at football or
3 cricket, they would have no inclination to meet on the battlefield. Even if one didn't know from
concrete examples (the 1936 Olympic Games, for instance) that international sporting contests
5 lead to hatred, one could deduce it from general principles.

Nearly all the sports practiced nowadays are competitive. You play to win, and the game
7 has little meaning unless you do your utmost to win. On the village green, where you pick up
sides and no feeling of local patriotism is involved, it is possible to play simply for the fun and
9 exercise; but as soon as the question of prestige arises, as soon as you feel that you and some
larger unit will be disgraced if you lose, the most savage combative instincts are aroused.
11 Anyone who has played even in a school football match knows this. At the international level
sport is frankly mimic warfare. But the significant thing is not the behavior of the players but the
13 attitude of the spectators and behind the spectators, of the nations who work themselves into
furies over these absurd contests, and seriously believe—at any rate for short periods—that
15 running, jumping and kicking a ball are tests of national virtue.

Questions:

1. How is the author's feeling when he heard about the benefit of sport?
2. What makes different nations have goodwill?
3. What is the example of international sport contest given by the author?
4. What is the suggestion for the nations to avoid battle?
5. What is the main feature of sport today?
6. What is the objective of doing sport at village?
7. What happens if we don't win?
8. Why do we feel disgraced if we lose?
9. What may be similar to combat or war?
10. What does the author suggest us to be aware of in any sport match?

Attachment 4.1.2: REVISED TEST ITEMS

| Feedbacks from the reviewer | Test item draft | Revised test items |
|--|---|--|
| Space for testee's identity is needed | No space given | Name: NIM : |
| Time allotment should be mentioned | No time allotment | Time: 60 minutes |
| Directions position | Written before the text | Written before the text and the questions |
| To clarify the second direction | 2. Answer the questions in brief (not more than 3 words) or simply by giving a Noun Phrase as in the following example: | Answer these questions based on the stated or implied information in the text. Write the answer in brief in the space provided (not necessarily in the form of full sentence). |
| To revise the wording of the example | <i>What is the text mainly discuss? (Answer: sporting spirit)</i> | <i>E.g. What is the author's mainpoint shown in the text? (Answer: sporting spirit)</i> |
| Give space to avoid wordy answers or testee's tendency to copy the whole sentence from the text | No space for the answer of each item | Given space for the answer of each item |
| No. 1 should be more general, replaced by no. 2 with clarification 'according to some people' | 1. How is the author's feeling when he heard about the benefit of sport? | 1. According to some people, what helps establish goodwill between nations? |
| The question no. 7 should be changed into more interpretative as its purpose is eliciting implicit information | 7. What happens if we don't win? | 7. How is doing sport at village different from doing sport in a larger unit? |
| Change the wording of no. 8 into a more understandable one | 8. Why do we feel disgraced if we lose? | 1. What might make us feel disgraced when we lose? |

Name:

NIM:

TEST OF READING II (Interpretive and Affective Reading)

Time: 60 minutes

Direction: Read the following passage carefully.

THE SPORTING SPIRIT

1 I am always amazed when I hear people saying that sport creates goodwill between the
nations, and that if only the common peoples of the world could meet one another at football or
3 cricket, they would have no inclination to meet on the battlefield. Even if one didn't know from
concrete examples (the 1936 Olympic Games, for instance) that international sporting contests
5 lead to hatred, one could deduce it from general principles.

Nearly all the sports practiced nowadays are competitive. You play to win, and the game
7 has little meaning unless you do your utmost to win. On the village green, where you pick up
sides and no feeling of local patriotism is involved, it is possible to play simply for the fun and
9 exercise; but as soon as the question of prestige arises, as soon as you feel that you and some
larger unit will be disgraced if you lose, the most savage combative instincts are aroused.
11 Anyone who has played even in a school football match knows this. At the international level
sport is frankly mimic warfare. But the significant thing is not the behavior of the players but the
13 attitude of the spectators and behind the spectators, of the nations who work themselves into
furies over these absurd contests, and seriously believe –at any rate for short periods—that
15 running, jumping and kicking a ball are tests of national virtue.

Direction: Answer these questions based on the stated or implied information in the text.

Write the answer in brief in the space provided (not necessarily in the form of full sentence).

E.g. What is the author's mainpoint shown in the text? (Answer: sporting spirit)

1. According to some people, what helps establish goodwill between nations?
.....
2. How is the author's feeling when he heard about the benefit of sport?
.....
3. What is the example of international sport contest given by the author?
.....
4. What is the suggestion for the nations to avoid battle?
.....
5. What is the main feature of sport today?
.....
6. What is the objective of doing sport at village?
.....
7. How is doing sport at village different from doing sport in a larger unit?
.....
8. What might make us feel disgraced when we lose?
.....
9. What may be similar to combat or war?
.....
10. What does the author suggest us to be aware of in any sport match?
.....

Attachment 4.3:

THE RELIABILITY FORMULA

Pearson Product-Moment Coefficient Correlation Formula:

$$r_{xy} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{N S_x S_y}$$

Notes:

r_{xy} : Coefficient correlation Pearson Product-Moment

\sum : sum

X : each student's score given by the first scorer

Y : each student's score given by the second scorer

Xbar : average of students' score given by the first scorer

Ybar : average of students' score given by the second scorer

N : the number of the students taking the test

Sx : Standard deviation of the test scores of the first scorer

Sy : Standard deviation of the test scores of the second scorer

Adapted from:

Djiwandono, M.S. 2008. *Tes Bahasa: Pegangan bagi Pengajar Bahasa. Jakarta: Indeks.*

Page: 173

SUMMARY OF NOTES OF TRYOUT TEST

| No. | Topic | Problems | What Needs to be Done |
|-----|---|--|--|
| 1. | Questions clarification | The text implies two views: the author's and some people's assumption | Question no. 4: "What is the suggestion for the nations to avoid battle?" is revised into: "What is the <u>author's</u> suggestion for the nations to avoid battle?" |
| 2. | Student's comprehension on the test direction | Many students tend to ignore the direction stating that the answers are based on explicit or implicit information from the text. They make inference to answer each question although the answer is explicitly stated in the text. | The teacher needs to give feedback on the following meeting to improve the student's comprehension on how to distinguish between explicit and implicit information in the text |
| 3. | Text wording | The original text uses informal wording which may confuse the testees such as: <ul style="list-style-type: none"> - Peoples - For the fun and exercise | The text wording should be changed into a standar one: <ul style="list-style-type: none"> - People - For fun and exercise |

TABLE OF SCORES**a. Scoring of Rater 1**

| Testee No. | Content | | | | | | | | | | Language | | | | | | | | | | Score |
|------------|---------|---|---|---|---|---|---|---|---|----|----------|---|---|---|---|---|---|---|---|----|-------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| 1. | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 3 | 3 | 2 | 3 | 3 | 2 | 56 |
| 2. | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 3 | 2 | 56 |
| 3. | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 2 | 2 | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 2 | 1 | 51 |
| 4. | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 1 | 2 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 2 | 2 | 51 |
| 5. | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 1 | 1 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 2 | 2 | 51 |
| 6. | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 2 | 2 | 2 | 3 | 2 | 2 | 3 | 2 | 50 |
| 7. | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 1 | 3 | 3 | 3 | 3 | 2 | 3 | 2 | 2 | 2 | 1 | 50 |
| 8. | 3 | 3 | 3 | 1 | 3 | 3 | 3 | 3 | 1 | 2 | 2 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 1 | 2 | 50 |
| 9. | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 1 | 1 | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 2 | 2 | 50 |
| 10. | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 1 | 2 | 3 | 3 | 3 | 2 | 2 | 3 | 3 | 2 | 2 | 1 | 49 |
| 11. | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 1 | 3 | 1 | 2 | 3 | 3 | 2 | 3 | 3 | 2 | 1 | 3 | 2 | 49 |
| 12. | 3 | 3 | 3 | 1 | 3 | 3 | 2 | 3 | 1 | 1 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 2 | 2 | 49 |
| 13. | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 1 | 1 | 2 | 2 | 3 | 2 | 3 | 3 | 3 | 3 | 1 | 2 | 49 |
| 14. | 3 | 3 | 3 | 1 | 3 | 3 | 3 | 3 | 1 | 2 | 3 | 3 | 3 | 1 | 3 | 3 | 3 | 3 | 1 | 2 | 48 |
| 15. | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 2 | 1 | 2 | 3 | 3 | 3 | 3 | 2 | 2 | 1 | 1 | 2 | 47 |
| 16. | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 1 | 1 | 3 | 2 | 2 | 1 | 3 | 3 | 3 | 3 | 1 | 1 | 46 |
| 17. | 3 | 3 | 3 | 3 | 1 | 1 | 3 | 3 | 1 | 2 | 2 | 3 | 2 | 2 | 1 | 1 | 3 | 3 | 2 | 2 | 44 |
| 18. | 3 | 3 | 3 | 2 | 3 | 3 | 2 | 1 | 1 | 1 | 3 | 2 | 3 | 1 | 3 | 3 | 3 | 1 | 2 | 1 | 44 |
| 19. | 3 | 3 | 3 | 2 | 1 | 3 | 3 | 3 | 1 | 1 | 1 | 3 | 3 | 1 | 1 | 3 | 3 | 3 | 1 | 1 | 43 |
| 20. | 3 | 3 | 3 | 2 | 1 | 3 | 2 | 1 | 1 | 2 | 3 | 3 | 3 | 1 | 1 | 3 | 2 | 1 | 1 | 2 | 41 |
| 21. | 3 | 1 | 3 | 1 | 1 | 3 | 3 | 3 | 1 | 1 | 3 | 1 | 3 | 3 | 1 | 3 | 2 | 1 | 1 | 2 | 40 |
| 22. | 1 | 3 | 3 | 1 | 3 | 2 | 1 | 3 | 1 | 2 | 1 | 1 | 3 | 1 | 3 | 1 | 1 | 3 | 1 | 2 | 37 |
| 23. | 3 | 1 | 3 | 1 | 1 | 3 | 3 | 1 | 1 | 1 | 2 | 3 | 2 | 2 | 1 | 3 | 2 | 2 | 1 | 1 | 37 |
| 24. | 3 | 3 | 3 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 33 |
| 25. | 2 | 3 | 2 | 2 | 1 | 3 | 3 | 1 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 2 | 3 | 1 | 1 | 1 | 33 |

b. Scoring of Rater 2

| Testee No. | Content | | | | | | | | | | Language | | | | | | | | | | Score |
|------------|---------|---|---|---|---|---|---|---|---|----|----------|---|---|---|---|---|---|---|---|----|-------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| 1. | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 2 | 57 |
| 2. | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 2 | 57 |
| 3. | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 2 | 2 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 1 | 52 |
| 4. | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 1 | 2 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 2 | 2 | 51 |
| 5. | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 1 | 1 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 2 | 2 | 51 |
| 6. | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 2 | 2 | 2 | 3 | 2 | 2 | 3 | 2 | 50 |
| 7. | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 1 | 3 | 3 | 3 | 3 | 2 | 3 | 2 | 2 | 2 | 1 | 50 |

| | | | | | | | | | | | | | | | | | | | | | |
|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|
| 8. | 3 | 3 | 3 | 1 | 3 | 3 | 3 | 3 | 1 | 2 | 2 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 1 | 2 | 50 |
| 9. | 3 | 3 | 2 | 1 | 3 | 3 | 3 | 3 | 1 | 1 | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 2 | 2 | 49 |
| 10. | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 1 | 2 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 2 | 2 | 1 | 50 |
| 11. | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 1 | 3 | 1 | 2 | 3 | 3 | 3 | 3 | 3 | 2 | 1 | 3 | 2 | 50 |
| 12. | 3 | 3 | 3 | 1 | 3 | 3 | 2 | 3 | 1 | 1 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 2 | 2 | 49 |
| 13. | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 1 | 1 | 2 | 2 | 3 | 2 | 3 | 3 | 2 | 3 | 1 | 2 | 48 |
| 14. | 3 | 3 | 3 | 1 | 3 | 3 | 3 | 3 | 1 | 2 | 3 | 3 | 3 | 1 | 3 | 3 | 2 | 3 | 1 | 2 | 47 |
| 15. | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 2 | 1 | 2 | 3 | 3 | 3 | 3 | 2 | 2 | 1 | 1 | 2 | 47 |
| 16. | 3 | 3 | 2 | 1 | 3 | 3 | 3 | 3 | 1 | 1 | 3 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 1 | 1 | 46 |
| 17. | 3 | 3 | 3 | 3 | 1 | 1 | 3 | 3 | 1 | 2 | 2 | 3 | 2 | 2 | 1 | 1 | 3 | 3 | 2 | 2 | 44 |
| 18. | 3 | 3 | 3 | 1 | 3 | 3 | 2 | 1 | 1 | 1 | 3 | 2 | 3 | 1 | 3 | 3 | 3 | 1 | 2 | 1 | 43 |
| 19. | 3 | 3 | 3 | 2 | 1 | 3 | 3 | 3 | 1 | 1 | 1 | 3 | 3 | 1 | 1 | 3 | 2 | 3 | 1 | 1 | 42 |
| 20. | 3 | 3 | 3 | 1 | 1 | 3 | 2 | 1 | 1 | 2 | 3 | 3 | 3 | 1 | 1 | 3 | 2 | 1 | 1 | 2 | 40 |
| 21. | 3 | 1 | 3 | 1 | 1 | 3 | 3 | 3 | 1 | 1 | 3 | 1 | 3 | 3 | 1 | 3 | 2 | 1 | 1 | 2 | 40 |
| 22. | 1 | 3 | 3 | 1 | 3 | 2 | 1 | 3 | 1 | 2 | 1 | 1 | 3 | 1 | 3 | 1 | 1 | 3 | 1 | 2 | 37 |
| 23. | 3 | 1 | 3 | 1 | 1 | 3 | 3 | 1 | 1 | 1 | 2 | 3 | 2 | 2 | 1 | 3 | 2 | 2 | 1 | 1 | 37 |
| 24. | 3 | 3 | 3 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 1 | 1 | 3 | 2 | 1 | 1 | 34 |
| 25. | 2 | 3 | 2 | 2 | 1 | 3 | 3 | 1 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 32 |

c. Average scores

| Testee No. | SCORES of RATER 1 | SCORES of RATER 2 | AVERAGE SCORES |
|------------|-------------------|-------------------|----------------|
| 1. | 56 | 57 | 56.5 |
| 2. | 56 | 57 | 56.5 |
| 3. | 51 | 52 | 51.5 |
| 4. | 51 | 51 | 51 |
| 5. | 51 | 51 | 51 |
| 6. | 50 | 50 | 50 |
| 7. | 50 | 50 | 50 |
| 8. | 50 | 50 | 50 |
| 9. | 50 | 49 | 49.5 |
| 10. | 49 | 50 | 49.5 |
| 11. | 49 | 50 | 49.5 |
| 12. | 49 | 49 | 49 |
| 13. | 49 | 48 | 48.5 |
| 14. | 48 | 47 | 47.5 |
| 15. | 47 | 47 | 47 |
| 16. | 46 | 46 | 46 |
| 17. | 44 | 44 | 44 |
| 18. | 44 | 43 | 43.5 |
| 19. | 43 | 42 | 42.5 |
| 20. | 41 | 40 | 40.5 |
| 21. | 40 | 40 | 40 |
| 22. | 37 | 37 | 37 |
| 23. | 37 | 37 | 37 |

| | | | |
|-----|----|----|------|
| 24. | 33 | 34 | 33.5 |
| 25. | 33 | 32 | 32.5 |

d. Simple Frequency Distribution

| SCORES | TALLY | SUM |
|---------------------|--------------------|-----|
| 56.5 | // | 2 |
| 51.5 | / | 1 |
| 51 | // | 2 |
| 50 | //// | 4 |
| 49.5 | // | 2 |
| 49 | / | 1 |
| 48.5 | / | 1 |
| 47.5 | / | 1 |
| 47 | / | 1 |
| 46 | / | 1 |
| 44 | / | 1 |
| 43.5 | / | 1 |
| 42.5 | / | 1 |
| 40.5 | / | 1 |
| 40 | / | 1 |
| 37 | // | 2 |
| 33.5 | / | 1 |
| 32.5 | / | 1 |
| Total scores = 1154 | Total testees = 25 | |

Attachment 5.2.2:

SIMPLE STATISTICS

| ASPECTS | SYMBOLS | FIGURES |
|--|---------------------|---------------------|
| Number of Testees | N | 25 |
| Number of Test Items | K | 10 |
| Central Tendencies: | — | |
| Mean | \bar{X} | 46.16 |
| Median | Med | 48.5 |
| Mode | Mod | 50 |
| Dispersion: | | |
| Standard Deviation | S | 6.3 |
| Variance | S^2 δ^2 | 39.85 |
| Range | R | 24 |
| Maximum score | Max | 56.5 |
| Minimum score | Min | 32.5 |
| Difficulty level | P | (see table 5.2.2.b) |
| Reliability (Pearson Product-Moment Correlation Coefficient) | r_{xy} | 0.99 |

a. Standard Deviation (SD)

| TESTEE NO. | SCORE (X) | AVERAGE SCORE (\bar{x}) | ($X - \bar{x}$) | ($X - \bar{x}$) ² |
|------------|-----------|-----------------------------|-------------------|--------------------------------|
| 1. | 56.5 | 46.16 | 10.36 | 107.33 |
| 2. | 56.5 | 46.16 | 10.36 | 107.33 |
| 3. | 51.5 | 46.16 | 5.36 | 28.73 |
| 4. | 51 | 46.16 | 4.86 | 23.62 |
| 5. | 51 | 46.16 | 4.86 | 23.62 |
| 6. | 50 | 46.16 | 3.86 | 14.90 |
| 7. | 50 | 46.16 | 3.86 | 14.90 |
| 8. | 50 | 46.16 | 3.86 | 14.90 |
| 9. | 49.5 | 46.16 | 3.36 | 11.29 |
| 10. | 49.5 | 46.16 | 3.36 | 11.29 |
| 11. | 49.5 | 46.16 | 3.36 | 11.29 |
| 12. | 49 | 46.16 | 2.86 | 8.18 |
| 13. | 48.5 | 46.16 | 2.36 | 5.57 |
| 14. | 47.5 | 46.16 | 1.36 | 1.85 |
| 15. | 47 | 46.16 | 0.86 | 0.74 |
| 16. | 46 | 46.16 | -0.14 | 0.02 |
| 17. | 44 | 46.16 | -2.14 | 4.58 |
| 18. | 43.5 | 46.16 | -2.64 | 6.97 |

| | | | | |
|------------------------|------|------------------------|----------------|--------|
| 19. | 42.5 | 46.16 | -3.64 | 13.25 |
| 20. | 40.5 | 46.16 | -5.64 | 31.81 |
| 21. | 40 | 46.16 | -6.14 | 37.70 |
| 22. | 37 | 46.16 | -9.14 | 83.54 |
| 23. | 37 | 46.16 | -9.14 | 83.54 |
| 24. | 33.5 | 46.16 | -12.64 | 159.77 |
| 25. | 32.5 | 46.16 | -13.64 | 186.05 |
| $\sum (X - \bar{x})^2$ | | | | 992.76 |
| N | | | | 25 |
| S | | $\sqrt{(992.76 : 25)}$ | $\sqrt{39.71}$ | 6.3 |

Reference: Djiwandono, M.S. 2008. *Tes Bahasa: Pegangan bagi Pengajar Bahasa. Jakarta: Indeks.* Page: 216

b. Difficulty level (p)

| Item | No. of correct answer | Level of difficulty (p) | Interpretation |
|------|-----------------------|-----------------------------|------------------|
| 1 | 23 | 0.92 | Easy |
| 2 | 23 | 0.92 | Easy |
| 3 | 21 | 0.84 | Fairly Easy |
| 4 | 8 | 0.32 | Fairly Difficult |
| 5 | 18 | 0.72 | Moderate |
| 6 | 22 | 0.88 | Fairly Easy |
| 7 | 20 | 0.80 | Fairly Easy |
| 8 | 10 | 0.40 | Moderate |
| 9 | 4 | 0.16 | Difficult |
| 10 | 1 | 0.04 | Very Difficult |

Formula: $p = \frac{\text{number of correct responses}}{\text{Total number of students}}$

Easy : 0.9
Moderate : 0.5
Difficult : 0.1

Reference: Djiwandono, M.S. 2008. *Tes Bahasa: Pegangan bagi Pengajar Bahasa. Jakarta: Indeks.* Page: 219

c. Reliability (Pearson Product-Moment Correlation Coefficient)

| Testee no. | Score X | Mean X (Mx) | (X-Mx) | (X-Mx) ² | Score Y | Mean Y (My) | (Y-My) | (Y-My) ² | (X-Mx)(Y-My) |
|---|---------|-------------|--------|---------------------|--|-------------|--------|---------------------|--------------|
| 1. | 56 | 46.16 | 9.84 | 118.37 | 57 | 46.12 | 10.88 | 107.06 | 107.06 |
| 2. | 56 | 46.16 | 9.84 | 118.37 | 57 | 46.12 | 10.88 | 107.06 | 107.06 |
| 3. | 51 | 46.16 | 4.84 | 34.57 | 52 | 46.12 | 5.88 | 28.46 | 28.46 |
| 4. | 51 | 46.16 | 4.84 | 23.81 | 51 | 46.12 | 4.88 | 23.62 | 23.62 |
| 5. | 51 | 46.16 | 4.84 | 23.81 | 51 | 46.12 | 4.88 | 23.62 | 23.62 |
| 6. | 50 | 46.16 | 3.84 | 15.05 | 50 | 46.12 | 3.88 | 14.90 | 14.90 |
| 7. | 50 | 46.16 | 3.84 | 15.05 | 50 | 46.12 | 3.88 | 14.90 | 14.90 |
| 8. | 50 | 46.16 | 3.84 | 15.05 | 50 | 46.12 | 3.88 | 14.90 | 14.90 |
| 9. | 50 | 46.16 | 3.84 | 8.29 | 49 | 46.12 | 2.88 | 11.06 | 11.06 |
| 10. | 49 | 46.16 | 2.84 | 15.05 | 50 | 46.12 | 3.88 | 11.02 | 11.02 |
| 11. | 49 | 46.16 | 2.84 | 15.05 | 50 | 46.12 | 3.88 | 11.02 | 11.02 |
| 12. | 49 | 46.16 | 2.84 | 8.29 | 49 | 46.12 | 2.88 | 8.18 | 8.18 |
| 13. | 49 | 46.16 | 2.84 | 3.53 | 48 | 46.12 | 1.88 | 5.34 | 5.34 |
| 14. | 48 | 46.16 | 1.84 | 0.77 | 47 | 46.12 | 0.88 | 1.62 | 1.62 |
| 15. | 47 | 46.16 | 0.84 | 0.77 | 47 | 46.12 | 0.88 | 0.74 | 0.74 |
| 16. | 46 | 46.16 | -0.16 | 0.01 | 46 | 46.12 | -0.12 | 0.02 | 0.02 |
| 17. | 44 | 46.16 | -2.16 | 4.49 | 44 | 46.12 | -2.12 | 4.58 | 4.58 |
| 18. | 44 | 46.16 | -2.16 | 9.73 | 43 | 46.12 | -3.12 | 6.74 | 6.74 |
| 19. | 43 | 46.16 | -3.16 | 16.97 | 42 | 46.12 | -4.12 | 13.02 | 13.02 |
| 20. | 41 | 46.16 | -5.16 | 37.45 | 40 | 46.12 | -6.12 | 31.58 | 31.58 |
| 21. | 40 | 46.16 | -6.16 | 37.45 | 40 | 46.12 | -6.12 | 37.70 | 37.70 |
| 22. | 37 | 46.16 | -9.16 | 83.17 | 37 | 46.12 | -9.12 | 83.54 | 83.54 |
| 23. | 37 | 46.16 | -9.16 | 83.17 | 37 | 46.12 | -9.12 | 83.54 | 83.54 |
| 24. | 33 | 46.16 | -13.16 | 146.89 | 34 | 46.12 | -12.12 | 159.50 | 159.50 |
| 25. | 33 | 46.16 | -13.16 | 199.37 | 32 | 46.12 | -14.12 | 185.82 | 185.82 |
| $\sum(X-Mx)^2 = 957.36$ | | | | | $\sum(Y-My)^2 = 1034.64$ | | | | |
| N = 25 | | | | | N = 25 | | | | |
| $S_x = \sqrt{957.36 : 25} = \sqrt{38.29} = 6.2$ | | | | | $S_y = \sqrt{1034.64 : 25} = \sqrt{41.39} = 6.4$ | | | | |
| $\sum (X-Mx)(Y-My) = 989.52$ | | | | | | | | | |
| N Sx Sy = 992 | | | | | | | | | |
| $r_{xy} = 989.52 : 992 = 0.99$ | | | | | | | | | |

Reference: Djiwandono, M.S. 2008. *Tes Bahasa: Pegangan bagi Pengajar Bahasa*. Jakarta: Indeks. Page: 174

Attachment 5.2.3:

TEST ITEM ANALYSIS

| ITEM | H (6) | M (13) | L (6) | H+M+L (25) | <i>P</i> | H – L | D | Quality |
|------|----------|-----------|----------|---------------|----------|-------|------|----------------------|
| 1 | 6 | 13 | 4 | 23 | 0.92 | 2 | 0.33 | Good enough |
| 2 | 6 | 13 | 4 | 23 | 0.92 | 2 | 0.33 | Good enough |
| 3 | 5 | 11 | 5 | 21 | 0.84 | 0 | 0.00 | Bad, need discarding |
| 4 | 4 | 4 | 0 | 8 | 0.32 | 4 | 0.67 | Very good |
| 5 | 6 | 11 | 1 | 18 | 0.72 | 5 | 0.83 | Very good |
| 6 | 6 | 12 | 4 | 22 | 0.88 | 2 | 0.33 | Good enough |
| 7 | 6 | 11 | 3 | 20 | 0.80 | 3 | 0.50 | Very good |
| 8 | 5 | 3 | 2 | 10 | 0.40 | 2 | 0.33 | Good enough |
| 9 | 4 | 0 | 0 | 4 | 0.16 | 4 | 0.67 | Very good |
| 10 | 0 | 1 | 0 | 1 | 0.04 | 0 | 0.00 | Bad, need discarding |

Notes:

N = 25

25% of N = 6

H (high achiever) and L (low achiever) = 6 each

p = difficulty level = (sum of correct responses) : (number of testees) = (H+M+L) : N

D = Discrimination power = (H – L) : (number of H or L)

D ≥ 0.40 : very good

D : 0.30 – 0.39 : good enough

D : 0.20 - 0.29 : not good, need revising

D ≤ 0.19 : bad, need discarding

Adapted from:

Djiwandono, M.S. 2008. *Tes Bahasa: Pegangan bagi Pengajar Bahasa. Jakarta: Indeks.*

Page: 224

Attachment 5.3:

THE GRADES

| NO. | STUDENTS NO. | SCORES | ROUND-UP SCORES | GRADES | PASS/NOT PASS |
|------------|---------------------|---------------|----------------------------|---------------|--------------------------|
| 1. | 309222416889 | 56.5 | 57 | A | P |
| 2. | 309222416893 | 56.5 | 57 | A | P |
| 3. | 109221426119 | 51.5 | 52 | A | P |
| 4. | 309222416885 | 51 | 51 | A | P |
| 5. | 309222426420 | 51 | 51 | A | P |
| 6. | 109221422438 | 50 | 50 | B | P |
| 7. | 309222422864 | 50 | 50 | B | P |
| 8. | 309222416887 | 50 | 50 | B | P |
| 9. | 209221423198 | 49.5 | 50 | B | P |
| 10. | 309222426124 | 49.5 | 50 | B | P |
| 11. | 209222426192 | 49.5 | 50 | B | P |
| 12. | 109221426105 | 49 | 49 | B | P |
| 13. | 309222426416 | 48.5 | 49 | B | P |
| 14. | 309222416886 | 47.5 | 48 | B | P |
| 15. | 209221416639 | 47 | 47 | B | P |
| 16. | 109221422112 | 46 | 46 | B | P |
| 17. | 309222416857 | 44 | 44 | B | P |
| 18. | 309222426423 | 43.5 | 44 | B | P |
| 19. | 309222426865 | 42.5 | 43 | B | P |
| 20. | 309222416888 | 40.5 | 41 | C | P |
| 21. | 109221426114 | 40 | 40 | C | P |
| 22. | 309222426418 | 37 | 37 | D | NP |
| 23. | 309222426417 | 37 | 37 | D | NP |
| 24. | 309222426422 | 33.5 | 34 | D | NP |
| 25. | 309222426863 | 32.5 | 33 | D | NP |

Note:

The grading is based on attachment 3.4

SUMMARY OF PROBLEMS AND HOW TO IMPROVE

| No. | Topic, location | What's wrong | How to improve |
|------------|---|--|--|
| 1. | Text wording | The original text uses informal wording which may confuse the testees such as: <ul style="list-style-type: none"> - Peoples - For the fun and exercise | The text wording should be changed into a standard one: <ul style="list-style-type: none"> - People - For fun and exercise |
| 2. | Level of difficulty | 2 items are too difficult | No. 3 and 10 need discarding or revised |
| 3. | Item analysis | 2 items are bad | No. 3 and 10 need discarding or revised |
| 4. | Questions clarification | The text implies two views: the author's and some people's assumption | Question no. 4: "What is the suggestion for the nations to avoid battle?" is revised into: "What is the <u>author's</u> suggestion for the nations to avoid battle?" |
| 5. | Student's comprehension on the test direction | Many students tend to ignore the direction stating that the answers are based on explicit or implicit information from the text. They make inference to answer each question although the answer is explicitly stated in the text. | The teacher needs to give feedback on the following meeting to improve the student's comprehension on how to distinguish between explicit and implicit information in the text |

SOME THEORETICAL JUSTIFICATION OF READING TEST

Introduction

In reading comprehension classes today, many reading skills and strategies for use in everyday situation have been put in various emphases. Unfortunately, attempts at dealing with the many complex reading skills frequently come too late, at university level (Heaton, 1989). Therefore, understanding better the kinds of skill to assess in reading comprehension is a significant start.

Assessing students' skill has become an inseparable part of language instructional process. By having a test, teacher can have fruitful information on learners' competence and achievement. Besides, it can serve as feedback provider for the previous process of English Language Teaching, the instructional objective formulation and instruction (Djiwandono, 1996). In line with this, conducting this **Test Development Project (TDP)** belongs to a crucial attempt to develop a good test through reviewing the instructional objective and understanding all about building a better test. The following discussion is some theoretical view of reading comprehension, testing reading and its analysis.

1. Reading

This section comprises the definition and nature of reading. The coverage of the discussion also includes the types of reading skills technique, reading model and reading stages. The understanding of these basic concepts of reading is very crucial to highlight before developing reading assessment instrument or TDP.

1.1 Definition of reading

Reading is an activity characterized by the translation of symbols, or letters, into words and sentences that have meaning to the individual. The ultimate goal of reading is to be able to understand written material, to evaluate it, and to use it for one's needs (Chall and Stahl, 2009). In this TDP the ultimate goal is to be able to understand the explicit and implicit information of the written material.

Reading, or known as reading comprehension has three levels namely literal, inferential and evaluational reading. **Literal reading** is comprehending text which refers to identifying written facts involving identification and recall of main ideas and supporting details, also

analysis and reorganization of ideas by summarizing. The **inferential reading** is comprehending text covering interpretive or interpreting implicit meaning, general inferential or interpreting the whole supporting details, and predictive or predicting the development of the content in the text. While, the **evaluational reading** is comprehending text covering judgemental by showing philosophical judgement on the author's agreement or disagreement, appreciative by giving emotional reactions to the text, and critical by questioning the information in the text (Barret in Indah, 2002). Among the three comprehension levels, this TDP deals with **interpretive reading**.

1.2 Nature of reading

Dealing with the nature of reading in second language setting, readers should use their background knowledge regarding the topic, text structure, their knowledge of the world, and their knowledge of print to interact with the text and to make predictions about it (O'Malley and Pierce, 1996).

Getting students to read English text concerns with many reasons. In the first place, students are given good exposure to English. Reading texts provide good models for English writing. In addition, reading texts also provide opportunities to study language: vocabulary, grammar, punctuation and the way to construct sentences, paragraphs and texts. Lastly, good reading texts can introduce interesting topics to stimulate discussion, excite imaginative responses and be the springboard for well-rounded, fascinating lessons (Harmer, 1998). These reasons become the consideration of this TDP developer in choosing the passage for the test.

Another consideration is also made due to reading principles. There are some basic principles of reading viewed from content, purpose and expectation, and receptive skills (Harmer, 1995). On the **content** aspect, reading can be done due to interest such as for enjoyment, pleasure or intellectual stimulation; and to usefulness like understanding manual or instruction. The **purpose and expectation** of reading deal with finding out or discovering certain information. Usually readers also have expectation about the content of the text before reading it. The **receptive skills** involved in reading cover six abilities namely predictive skills, extracting specific information, getting the general picture, extracting detailed information, recognizing function and discourse patterns, and deducing meaning from context.

There are some characteristics of efficient and inefficient reading underpinning this TDP. They are as listed by Ur (1996: 148) in the following:

| Aspect | <i>Efficient</i> | <i>Inefficient</i> |
|---------------------------------------|---|---|
| 1. Language | The language of the text is comprehensible to the learners | The language of the text is too difficult |
| 2. Content | The content of the text is accessible to the learners; they know enough about it to be able to apply their own background knowledge. | The text is too difficult in the sense that the content is too far removed from the knowledge and experience of the learners. |
| 3. Speed | The reading progresses fairly fast: mainly because the reader has 'automized' recognition of common combinations, and does not waste time working out each word or group of words anew. | The reading is slow: the reader does not have a large 'vocabulary' of automatically recognized items. |
| 4. Attention | The reader concentrates on the significant bits and skims the rest, may even skip parts she or she knows to be insignificant. | The reader pays the same amount of attention to all parts of the text. |
| 5. Incomprehensible vocabulary | The reader takes incomprehensible vocabulary in his or her stride: guesses its meaning from the surrounding text, or ignores it and manages without; uses a dictionary only when these strategies are insufficient. | The reader cannot tolerate incomprehensible vocabulary items; stops to look every one up in a dictionary; and/or feels discouraged from trying to comprehend the text as a whole. |
| 6. Prediction | The readers think ahead, hypothesizes, predicts. | The reader does not think ahead, deals with the text as it comes. |
| 7. Background information | The reader has and uses background information to help understand the text. | The reader does not have or use background information. |
| 8. Motivation | The reader is motivated to read: by interesting content or a challenging task. | The reader has no particular interest in reading. |
| 9. Purpose | The reader is aware of a clear purpose in reading; for example to find something, to get pleasure. | The reader has no clear purpose other than to obey the teacher's instruction. |
| 10. Strategies | The reader uses different strategies for different kinds of reading. | The reader uses the same strategy for all texts. |

1.3 Types of reading skills

In reading, there are four identifiable skills. *First*, **skimming** which is reading for gist, e.g. quickly glancing through an article to see if it interests the reader. *Second*, **scanning** or reading to locate specific information, e.g. locating a telephone number in a directory. *Third*, **intensive reading** where the reader is trying to absorb all the information given, e.g. reading dosage instruction for medicine. *Fourth*, **extensive reading** where the reader deals with a longer text as a whole, which requires the ability to understand the component parts and their contribution to the overall meaning, e.g. reading a newspaper article, short story or novel (Wright, 1992). Based on the syllabus referred in this TDP, the third skill namely intensive reading has been the emphasis.

Examining reading difficulties must be related with the **specific skills** to be measured. Heaton (1989) listed some specific skills involved in reading as the following:

- Recognize words and word groups
- Deduce the meaning of words by understanding word formation and contextual clues
- Understand explicitly stated information
- Understand relations within the sentence, especially elements of sentence structure, negation, fronting and theme, and complex embedding
- Understand relations between parts of text through lexical devices (e.g. repetition, synonyms), cohesive device (e.g. pronoun), and connectives (e.g. adverb)
- Perceive temporal and spatial relationship, and also sequences of ideas
- Understand conceptual meaning such as quantity, definiteness, comparison, cause, etc.
- Anticipate and predict what will come next in the text
- Identify the main idea
- Understand implicit information
- Look for general meaning and read for specific information
- Read critically, and
- Adopt a flexible approach and vary reading strategies according to the reading purpose.

Among the above skills, two main abilities belong to **reading micro-skills** (Wright, 1992) namely predicting and recognizing implied meaning. This is because in dealing with any texts, readers use their knowledge and understanding of the context to make predictions about what they think will happen or be read from the text. Not only does the context help readers to make prediction, it also allows them to assume what is being referred to by the words in the text. Whereas, in this TDP the skills underlined are **understanding explicit and implicit information**.

1.4 Reading model and stages

Based on its model, reading includes discovering meaning in print and script, within a social context, through bottom-up and top-down processing and the use of strategies and skills (Gebhard, 2000). The **bottom-up process** starts with the ability to recognize words, phrases and sentences, while the **top-down process** relies on the background knowledge related to the content of the text.

For second language learners, Eskey believes that top-down model does not appear to fit the process of reading unless the learners are already proficient learners (in O'Malley and Pierce, 1996). This model requires higher level skills, such as predicting meaning with contextual clues or background knowledge, at the expense of lower-level skills such as rapid and accurate identification of vocabulary and syntactical forms. The learners of this TDP might use the combination of the two processes. When dealing with the task to find the explicit information, the learners use bottom-up process. When asked to dig the implicit information, they would employ the top down process.

In addition to the reading model employed, Anderson and Pearson found that second language reading learners depend on specific examples in memory as well as on abstract and general schemata because word meanings are context sensitive (in O'Malley and Pierce, 1996). **Schemata** refers to knowledge already stored in memory, while abstract and general schemata refer to distinctive features that make up generic categories, such as *bird*, *bachelor* and *door*. Readers background knowledge is often culture-bound and may not match with the schemata needed for a given reading text. In this TDP, the students should have general knowledge or schemata on sport as the given text discusses about the sporting spirit, otherwise they may find it difficult if they know nothing about sport.

There are three foremost stages in reading: pre-reading, during reading and post-reading. **Pre-reading** work should involve the student, create a sense of purpose in reading and help to focus the student's mind on the context of the text. It prepares the student for the content of the text and give them a particular task to do. **Post-reading** task helps to show what the students have understood and how they have responded to it (O'Malley and Pierce, 1996). Based on the teacher's information, these reading stages are mainly employed in the teaching and learning process in the class used in this TDP.

In addition to the distinction above, particularly in reading comprehension there are three stages namely **knowing** which includes recognizing sentence structure and recalling factual information, **comprehending** which refers to finding meaning and paraphrasing sentences, and **applying** which aims at being able form generalization and drawing conclusions (Oka, 1992). In doing the test developed in this TDP, the students involve the three stages above all at once.

The stages in reading may involve various activities. In the following Ur (1996: 146) suggests some ideas for **reading activities** some of which have been practised in the class of this TDP:

- Pre-question. A general question given before reading, asking the learners to find out a piece of information central to the understanding of the text.
- Do-it-yourself questions. Learners compose and answer their own questions
- Provide a title. Learners suggest a title if none was given originally, or an alternative, if there was.
- Summarize. Learners summarize the content in a sentence or two.
- Continue. The text is a story, learners suggest what might happen next.
- Preface. The text is a story, learners suggest what might have happened before.
- Gapped text. Toward the end of the text, four or five gaps are left that can only be filled in if the text has been understood. It is different from the conventional cloze-test (a text with regular gaps throughout) which tests grammatical and lexical accuracy and actually discourages purposeful, fluent reading.
- Mistakes in the text. The text has occasional mistakes. Learners are told in advance how many mistakes to look for.

- Comparison. There are two texts on a similar topic, learners note points of similarity or difference of content.
- Responding. The text is a letter or a provocative article, learners discuss how they would respond, or write an answer.
- Re-presentation of content. The text gives information or tells a story, learners re-present its content through a different graphic medium, e.g. map, diagram, etc.

Regarding the reading activities above, there are some **teaching reading principles** which are crucial to be noted by the test developer of this TDP. (1) Reading is not merely a passive skill. Readers have to understand what the words mean, understand the arguments, and work out if there is disagreement. (2) Students need to be engaged with what they are reading. When they are fired up by the topic or the task, they get much more from what is in front of them. (3) Students should be encouraged to respond to the content of a reading text, not just to the language. They should be allowed to express their feeling about the topic. (4) Prediction is a major factor in reading. By predicting, expectations are set up and the active process of reading is ready to begin. (5) Match the task to the topic. By choosing good reading tasks, the most commonplace passage can be made really exciting with imaginative and challenging tasks. (6) Good teachers exploit reading texts to the full. It means integrating the reading text into interesting class sequence using the topic for discussion and further tasks (Harmer, 1998).

2. Testing reading

In order to assess the reading skill which becomes useful in student evaluation, Routman (in O'Malley and Pierce, 1996) gives **suggestions for reading teachers** including the TDP developer to:

- Be thoroughly familiar with developmental learning process and curriculum
- Articulate a philosophy of assessment and evaluation
- Know about and have experience collecting, recording, interpreting and analyzing multiple sources of data
- Be flexible and willing to try out multiple assessment procedures, and
- Be committed to understanding and implementing an approach to evaluation that informs students and directs instruction.

2.1 What to test in reading

All language testing, including reading test, must be based on clear definition of language ability being measured.

2.2 Different ways of testing reading

Based on the student's level, the reading comprehension test should **measure certain abilities** as listed by Far (adapted in Djiwandono, 2008) as follow:

| No. | Level | Abilities |
|-----|---------------------|--|
| 1. | Elementary | <ol style="list-style-type: none">1) Understand the meaning of words used in text2) Identify the structure of the text and the relations between the paragraph3) Identify the main idea of the text4) Find the answer of the question on explicitly stated information in the text |
| 2. | Intermediate | <ol style="list-style-type: none">1) Understand the meaning of words used in text2) Identify the structure of the text and the relations between the paragraph3) Identify the main idea of the text4) Find the answer of the question on explicitly stated information in the text5) Answer the questions based on the text although the wording is dissimilar6) Make inference from the text content |
| 3. | Advanced | <ol style="list-style-type: none">1) Understand the meaning of words used in text2) Identify the structure of the text and the relations between the paragraph3) Identify the main idea of the text4) Find the answer of the question on explicitly stated information in the text5) Answer the questions based on the text although the wording is dissimilar6) Make inference from the text content7) Identify and understand the words and expression to comprehend literary nuance8) Identify and comprehend the writer's message as part of understanding the writer of the text |

The class used in this TDP belongs to pre-advanced level, the abilities covered the combination of intermediate and advanced level as mentioned above.

2.3 Reading test format

The idea of testing reading is to assess the ability to understand written ideas (Djiwandono, 1996). Basically, testing reading covers two format namely objective and subjective test. **Objective tests** involve: matching test including word matching, sentence matching, picture and sentence matching; true/false; and multiple choice. Whereas, **subjective test** could be in the form of Wh-questions; paraphrasing; and summarizing.

Eliciting completely subjective responses in reading test is also called 'open-ended' (Heaton, 1989). The response required may range from a one-word answer to one or two sentences:

- (One word answer): Where did the writer have a bad accident? Give the name of the town.
- (Answer in few words): You have a friend who is keen on cross-country running. Which event can he enter at the end of the month?
- (Sentence answer): According to the article, why do you think so few foreign cars have been imported into Singapore recently?

In this TDP, the second and the last responses are used instead of the first one.

2.4 Reading test item development

The discussion on test item development consists of two main points namely reading text selection and the writing of the reading test items.

2.4.1 Reading text selection

There at least four aspects to consider in **selecting reading test passage** in this TDP according to Harris (1969) as follow:

- a. Length.** There should be a sufficient content to yield and accordingly the passage should be kept brief. Passages between 100 and 250 words are about the proper length for intermediate level.
- b. Subject matter.** The excerpts or selected readings must be clear and meaningful when taken out of context and not require outside subject-matter information to be fully comprehended. On the other hand, the passages should not deal with information that is universally known, for in this case the testee may be able to answer correctly without paying much attention to the passage.

c. Style and treatment of subject. Reading tests should include materials of various types and styles. The literary excellence of a paragraph does not always mean suitable as a test passage. Paragraphs that make just one clear, direct point, for instance, seldom make suitable passages for testing purposes, inasmuch as they do not yield a sufficient number of test items.

d. Language. Test writer must select realistic task, therefore, passages that are overloaded with extremely difficult lexical items or complex syntactical structure may have to be adapted. The simplification must be carefully controlled so that the test does not fail to discriminate between various levels of reading proficiency.

In addition to the aspects above, Hughes (2004) mentions a number of parameters such as type, form, graphic features, topic, style, intended readership, length, readability or difficulty, range of vocabulary and grammatical structure.

There are two types of texts based on the design, **authentic text** and non-authentic text. The former are said to be those which are designed for native speakers: they are 'real' texts designed not for language students, but for the native speakers of the language in question, for instance English language newspaper and advertisement. A **non-authentic text** in language teaching terms is one that has been written especially for language students. Such texts sometimes concentrate on the language they wish to teach (Harmer: 1995). The latter is chosen as the text for this TDP.

2.4.2 Writing of the reading test items

The starting point of writing items is a careful reading of the text, having the specified objective listed. The next step is to decide what task it is reasonable to expect candidates to be able to perform in relation to the objectives. It is only then that drafts items should be written (Hughes, 2004). Paragraph numbers and line numbers should be added to the text. These steps are applied in this TDP.

After the test items are constructed, the next process is **validation**. The text and items should be presented to a reviewer using a moderation checklist adapted from Hughes (2004) as an example below:

| No. | Points to review | Yes | No |
|-----|---|-----|----|
| 1. | Is the English of text and items are grammatically correct? | | |
| 2. | Is the English natural and acceptable? | | |

| | | | |
|----|---|--|--|
| 3. | Is the item in accordance with specified objectives? | | |
| 4. | Is specified reading sub-skills necessary in order to respond correctly? | | |
| 5. | Short answer: Is answer within productive abilities? Can it be scored validly and reliably? | | |
| 6. | Is the item economical? | | |
| 7. | Is the key complete and correct? | | |

3. Test analysis

In analyzing the tryout test result, there are some concepts need to be reviewed such as scoring, grading, validity and reliability.

3.1 Scoring reading test

As mentioned in 2.3 on reading test format, the Wh-question items are used in this TDP. Therefore, the scoring awards at least two or three marks for each correct answer. If the maximum for a correct answer is three marks, the example given by Heaton (1989) is as follows:

- Correct answer in a grammatically correct sentence or a sentence containing only a minor error3
- Correct answer in a sentence containing one or two minor errors (but causing no difficulty in understanding)2
- Correct answer but very difficult to understand because of one or more major grammatical errors1
- Incorrect answer in a sentence with or without errors0

3.2 Grading reading test

Grading can be done by two ways namely absolute grading which is used in this TDP and relative grading. The first refers to the grades that can be described directly in terms of student performance of others (Gronlund and Waugh, 2009). The set of grades might be expressed as one of the following:

| Grades | Points | Points | Points |
|--------|----------|----------|----------|
| A | 90 – 100 | 95 – 100 | 91 – 100 |
| B | 80 – 89 | 85 – 94 | 86 – 90 |
| C | 70 – 79 | 75 – 84 | 81 – 85 |
| D | 60 – 69 | 65 – 74 | 75 – 80 |
| F | Below 60 | Below 65 | Below 75 |

In relative grading, students are typically ranked in order of performance (based on a set of test scores or combined assessment results). The proportion of students should receive each grade is predetermined and might appear as one of the following:

| Grades | Percent of students | Percent of students |
|--------|---------------------|---------------------|
| A | 15 | 10 – 20 |
| B | 25 | 20 – 30 |
| C | 45 | 40 – 50 |
| D | 10 | 10 – 20 |
| F | 5 | 0 – 10 |

In grading reading test, scores obtained from a test need to be interpreted by to make decision. There are two kinds of grading reference namely **Criterion reference** and **Norm reference**. The former, which is used in this TDP, is designed to measure well- defined and specific instructional objectives (or table of specification) that are specific to a particular program namely reading for understanding implicit and explicit information from the text. It depends on the well- established standard or criteria (minimum acceptable performance level). Whereas, Norm reference interpretes the scores relative to the scores of other students who took the test by comparing with the normal distribution. The purpose is to spread the students along the continuum of scores so that those with low ability are at one left end of the normal distribution, while those with high ability are at the other end. This is correlated with the relative grading as presented previously.

Another concern of the grading in this TDP is making effective and fair grading. Gronlund and Waugh (2009) suggest guidelines providing a framework that should help clarify and standardize the task as follow:

- Inform students at the beginning of instruction what grading procedures will be used.
- Base grades on student achievement, and achievement only.
- Base grades on a wide variety of valid assessment data.
- When combining scores for grading, use a proper weighting technique.
- Select an appropriate frame of reference for grading

3.3 The validity of reading test

A reading test is valid if it really measures what should be tested. Validity is the relevance of a certain test to the kind of competence or ability as the main object to measure. In reading it deals with the competence to understand the information in written text. There are some kinds of validity namely criterion, content, face and construct validity (Brown, 2001 & Djiwandono, 2008).

This TDP uses **content validity** so that the test has reflected samples of the subject matter about which conclusions are to be drawn. The content is derived from the textbook as the source of test material referring to the syllabus dealing with the ability to understand implicit and expliting information from the text.

Although the focus of content-related evidence of validity is on the adequacy of the sampling, a valid interpretation of the assessment results assumes that the assessment was properly prepared, administered, and scored. Validity can be lowered by inadequate procedures in any of areas (Gronlund and Waugh, 2009). Thus, validity is 'built in' during the planning and preparation stages and maintained by proper administration and scoring as what implemented in this TDP.

In order to judge whether or not a test has content validity, specification of the skills is needed and made early at very early stage to provide test developer with the basis for making a principled selection of elements for inclusion in the test (Hughes, 2004). The table of specification is as shown in the attachment of this TDP.

In maintaining the validity, the TDP developer considers the factors that lower the validity of assessment result. *First*, tasks that provide an inadequate sample of the achievement to be assessed. *Second*, tasks that do not function as intended, due to use of improper types of tasks, lack of relevance, ambiguity, clues, bias, inappropriate difficulty, or similar factors. *Third*, improper arrangement of tasks and unclear directions. *Fourth*, too few tasks for the types of interpretation to be made (e.g., interpretation by objective basedon a few test items). *Fifth*, improper administration, such as inadequate time allowed and poorly controlled conditions. And *sixth*, judgmental scoring that uses inadequate scoring guides, objective scoring that contains computational errors (Gronlund and Waugh, 2009).

3.4 The reliability of reading test

In general, the reliability of a test refers to the consistency of test results (Ebel and Frisbie, 1986). Unlike objective-type language tests, however, the major concern about the reliability of subjective test in reading is rater reliability because the scores are obtained from raters' judgement.

Conceptually, there are two types of rater reliability: **intra-rater reliability** and **inter-rater reliability**. The former refers to the consistency of a rater in judging the same paper at two different points of time. It is the consistency of the first and the second reading of the same rater upon the same paper. To obtain intra-rater reliability coefficient, a set of scores from the first and the second reading are correlated. The obtained correlation coefficient indicates the degree of the consistency of the rater in rating the same paper at different points of time.

Inter-rater reliability refers to the degree of a close agreement of two or more raters in judging the same paper. Similar to intra-rater reliability, to obtain inter-reliability coefficient, a set of scores from each rater are correlated. The obtained correlation coefficient indicates the degree of closeness of agreement between them in judging the same paper.

The try out analysis of this TDP applied inter-rater reliability as the papers were scored by two raters. The computation of the two sets of scores would show the correlation coefficient which ranges from 0.00 to 1.00. The closer to one, the higher the agreement. Intra-rater reliability computation was not applied as it is commonly used by a classroom teacher who assesses his/her own students for grading purposes.

There are some factors lowering the reliability of test scores which become the concern of this TDP reliability. *First*, test scores are based on too few items. In the remedy, longer tests or accumulate scores from several short tests can be used. *Second*, range of scores is too limited, therefore in the remedy the item difficulty need adjustment to obtain larger spread of scores. *Third*, testing conditions are inadequate, so that in the remedy the time is arranged and any interruption, noise or other disrupting factors are eliminated. And *fourth*, scoring is subjective therefore in the remedy scoring keys are prepared and followed carefully (Gronlund and Waugh, 2009).

In addition there are more ways taken into consideration by this TDP developer to make the reading test more reliable as suggested by Hughes (2004) as follows:

- Take enough samples of behaviour.

- Exclude items which do not discriminate well between weaker and stronger students.
- Do not allow testees to much freedom.
- Write unambiguous items.
- Provide clear and explicit instructions.
- Ensure that test is well laid out and perfectly legible.
- Make testees familiar with format and testing technique.
- Provide uniform and non-distracting conditions of administration.
- Use items that permit scoring which is as objective as possible.
- Make comparisons between testees as direct as possible.
- Provide a detailed scoring guide.
- Train scorers.
- Agree acceptable responses and appropriate scores at outset of scoring.
- Identify testees by number, not name.
- Employ multiple, independent scoring.

4. Lessons learned in developing reading test

The whole stages in this TDP become a significant learning process for the test developer. What the literature can do is only providing theoretical bases for developing reading test, but the real practice implemented in this one full semester TDP is much more fruitful. This is in line with the philosophy of learning: “Tell me and I’ll forget, show me and I’ll remember, and involve me and I’ll learn” (Confucious).

Implementing the first stage in TDP namely the **test preparation** has guided the test developer to make careful analysis in deciding the prospective test takers, institution and level. This was not easy since the test developer needed to make sure that this first step in TDP would facilitate the further process.

The second step, **test plan**, has trained the test developer intensively on how to employ critical thinking in analyzing the given syllabus to be compared with the references and proposed a new one by making justification.

The next step which is more demanding is **test development**. As the actual gist of the process, the developer learned a lot from the way how test items should be constructed

step by step. Started from determining the test content including the text to be chosen, developing scoring guide, setting up the criteria suitable with the table of specification and setting the scoring procedure. Setting up the criteria happened to be a great lesson for the test developer. Based on her teaching and testing experience using intra-rater checking, the scoring took longer time than by using the criteria employed in this TDP.

The **test validation** as the next process is also a significant aspect where the reviewer and the test developer learned from each other by discussing the test items' strength and weakness.

The last step is **test tryout**, became very crucial in finding out the real application of the test being constructed. The most interesting process is in doing the analysis of the tryout result namely examining the score result, having careful analysis on the different scores between the raters to decide the standard deviation and compute the reliability coefficient. Processing the item analysis and the item difficulty is also challenging to see whether the previous stage namely test development has worked well. Furthermore, by the grading, not only the quality of the students but also the quality of the test items can be obtained.

Constructing the **project report** is the ultimate accomplishment which requires critical comment, details of the attachment, and significant learning concept. Above all, this brings inspiring ideas for a possible research relevant to this TDP in the future.

5. List of reference

- Brown, H.D. 2001. *Teaching by Principles: An Interactive Approach to Language Pedagogy* (2nd Ed.). New York: Longman.
- Chall, J.S and Stahl, S. Microsoft Encarta Premium 2009. © 1993-2008 Microsoft Corporation
- Djiwandono, M.S. 2008. *Tes Bahasa: Pegangan bagi Pengajar Bahasa*. Jakarta: PT Indeks.
- Djiwandono, M.S. 1996. *Test Bahasa dalam Pengajaran*. Bandung: Penerbit ITB.
- Ebel, R.L. And Frisbie, D.A. 1986. *Essentials of Educational Measurement*. London: Prentice-Hall Inc.
- Gebhard, J.G. 2000. *Teaching English as a Foreign or Second Language: A Teacher Self-development and Methodology Guide*. Ann Arbor: The University of Michigan Press.
- Gronlund, N.E. and Waugh, C.K. 2009. *Assessment of Student Achievement* (Ninth Edition). New Jersey: Pearson Education, Inc.

- Harmer, J. 1998. *How to Teach English: An Introduction to the Practice of English Language Teaching*. Essex: Addison Wesley Longman Limited.
- Harmer, J. 1995. *The Practice of English Language Teaching*. New York: Longman Group
- Harris, D.P. 1969. *Testing English as a Second Language*. New York: McGraw-Hill.
- Heaton, J.B. 1989. *Writing English Language Tests*. London: Longman Group UK Limited.
- Hughes, A. 2004. *Testing for Language Teachers* (2nd edition). Cambridge: Cambridge University Press.
- Indah, R.N. 2002. *Enhancing Content-Area Reading Skills of The English Department Students of STAIN Malang through Summary Writing*. Unpublished Thesis. Malang: PPs UM.
- Oka, D.D. 1992. *Bahan Pengajaran (Readers) Reading II Semester 3*. Unpublished textbook. Malang: FPBS IKIP Malang.
- O'Malley, J.M. and Pierce, L.V. 1996. *Authentic Assessment for English Language Learners: Practical Approach for Teachers*. New York: Addison-Wesley Publishing Company, Inc.
- Ur, P. 1996. *A Course in Language Teaching: Practice and Theory*. Melbourne: Cambridge University Press.
- Wright, A. 1992. *Pictures for Language Learning*. Victoria: Cambridge University Press.